

# DNA fingerprinting on trial

Eric S. Lander

In the rush to use the tremendous power of DNA fingerprinting as a forensic tool, the need for standards has been overlooked.

WITH the exception of identical twins, no human beings have identical DNA sequences. Of the 3,000 million nucleotides which we inherit from each parent, about 1 in 1,000 is a site of variation, or polymorphism, in the population. These DNA polymorphisms are most conveniently detected when they alter the length of the DNA fragments produced by the action of restriction enzymes, giving rise to restriction fragment length polymorphisms (RFLPs). In standard practice, the length of the fragments is measured by the rate at which they move in an electrophoresis gel.

More than 3,000 RFLPs have been identified to date, including some 100 highly polymorphic loci at which dozens of variant alleles are present in the population. By using RFLPs to trace the inheritance of chromosomal regions in families afflicted with genetic disorders, human geneticists have been able to pinpoint the location of the genes causing diseases such as Huntington's disease, cystic fibrosis and others — in the process spawning the field of DNA diagnostics.

Forensic science has more recently latched onto RFLPs, but with a different purpose: to identify the individual origin of blood or semen samples found in criminal investigations based on their distinctive RFLP patterns. In the United States, forensic RFLP testing has been pioneered by two private laboratories — Lifecodes Corporation of Valhalla, New York and Cellmark Diagnostics of Germantown, Maryland — and is also used by the Federal Bureau of Investigation, which began testing earlier this year.

Since the first use of DNA 'fingerprinting' in a trial in Florida in 1988, DNA 'fingerprint' evidence has already been used in more than 80 criminal trials in the United States. Applying the legal standard for the admissibility of novel scientific evidence defined in *United States v. Frye* in 1923, trial judges have raced to admit DNA fingerprinting as evidence on the grounds that the methods are "generally accepted in the scientific community", citing the application of RFLPs in DNA diagnostics and accepting claims that false positives are virtually impossible.

With due respect, the courts have been too hasty. Although DNA fingerprinting clearly offers tremendous potential as a forensic tool, the rush to court has obscured two critical points: first, DNA fingerprinting is far more technically demanding

than DNA diagnostics; and second, the scientific community has not yet agreed on standards that ensure the reliability of the evidence.

DNA diagnostics requires simply identifying whether each parent has passed to a child the RFLP pattern inherited from his or her mother or father. Because the four discrete patterns are known in advance, these investigations have built-in consistency checks which guard against many errors and artefacts.

DNA fingerprinting, by contrast, is more like analytical biochemistry: one must determine whether two completely unknown samples are identical. Because hypervariable RFLP loci often involve 50–100 alleles yielding restriction fragments of very similar lengths, reliably recognizing a match is technically demanding. At one commonly used locus, for example, most alleles lie within a mere 2 per cent of the length of the gel.

Few molecular geneticists, in such circumstances, would declare a match without performing a mixing experiment, in which a 50:50 mixture of the two samples is shown to yield the same pattern as each sample separately. In the rare event that a mixing experiment could not be carried out, most molecular geneticists would at least insist on using internal controls — probes which detect non-polymorphic DNA fragments within each lane — to verify that the lanes have run at equal speeds and to provide standards against which fragment sizes can be measured precisely.

Yet the DNA fingerprinting results now being introduced into the US criminal courts are often based on much flimsier evidence. Not only are mixing experiments and internal controls often omitted, but some laboratories use no objective standards whatsoever for declaring a match.

Unlike DNA diagnostics, DNA fingerprinting also depends on inferences about the frequency with which matching RFLP patterns will be found by chance, which in turn rest on simplifying assumptions about population genetics whose accuracy has not yet been rigorously tested for highly polymorphic RFLP loci. For example, it is assumed without convincing proof that Caucasians, Blacks and Hispanics can each be regarded as homogeneously mixed populations, without significant subgroups, even when considering loci at which most alleles are relatively young from the perspectives of population genetics.

Yet despite such fundamental uncertainties, forensic laboratories blithely cite breathtaking frequencies: a recent report based on the study of only four RFLPs announced that the chance of an alleged match occurring at random was 1 in 738,000,000,000,000.

It is my belief that we, the scientific community, have failed to set rigorous standards to which which courts, attorneys and forensic-testing laboratories can look for guidance — with the result that some of the conclusions presented to courts are quite unreliable.

My concern is not merely academic: during the past five months, I have been an advisor for the defence and given six days of testimony in what has turned out to be the longest and most searching pre-trial Frye hearing in the United States on the admissibility of DNA evidence — a murder case in the Bronx, one of the boroughs of New York City. I have also had occasion to investigate several other DNA fingerprinting cases in the course of preparing a report at the request of the US Office of Technology Assessment, to be delivered later this summer. (My own field is medical genetics: both projects arose as unintended consequences of accepting an invitation to a conference on DNA forensics at Cold Spring Harbor's Banbury Center in late 1988.)

It is my contention that DNA forensics sorely lacks adequate guidelines for the *interpretation* of results — both in molecular biology and in population genetics. To illustrate this, I will draw examples from cases with which I am personally familiar. I should emphasize that the focus on specific cases is not intended to criticize particular testing laboratories: with scientific consensus lacking, similar disagreements about interpretation would surely arise in many other cases involving other laboratories.

## The Castro case

On 5 February 1987, Vilma Ponce and her 2-year-old daughter were stabbed to death in their Bronx apartment. Acting on a tip, police interrogated a neighbourhood handyman, Jose Castro. Detectives noticed a small bloodstain on Castro's watch, which was sent for analysis to Lifecodes. Company scientists extracted about 0.5 µg of DNA from the bloodstain, which they compared with DNA from the two victims.

The DNA was digested with the restric-

tion enzyme *Pst*1, size-fractionated on an agarose gel, and transferred onto a Southern blot. The blot was then hybridized with probes for three RFLP loci; DXYS14, D2S44 and D17S79, as well as a probe for a Y-chromosome locus to identify sex. (Human loci detected by random DNA probes are named according to chromosome and order of discovery.) On 22 July 1987, Lifecodes issued a formal report<sup>1</sup> to the district attorney (Table 1) stating that the DNA patterns on the watch and the mother matched, and reporting the frequency of the pattern to be about 1 in 100,000,000 in the Hispanic population. The report indicated no difficulties or ambiguities. Yet there are several fundamental difficulties, as follows:

**DXYS14: Identifying bands.** Contrary to the forensic report (Table 1), the only autoradiogram involving DXYS14 shows

who had provided the probe to Lifecodes, testified that the DXYS14 autoradiogram had to be considered to exclude the defendant in the absence of any experiments to explain away the non-matching bands.

Lifecodes' discounting of the two non-matching bands in the watch lane suggests that its identification of bands may have been influenced by making direct comparisons between lanes containing different DNA samples, rather than by considering each lane in its own right. Additional support for this hypothesis is provided by two further examples:

(1) In the dead daughter's pattern at DXYS14, all other expert witnesses for the prosecution and the defence identified only a single band (Fig. 1). However, Lifecodes' laboratory records show that it recorded three bands in this lane — in precisely the same positions as those recorded for the mother and the watch.

TABLE 1 Reported fragment sizes (in kilobases) at three RFLPs in *New York v. Castro*, as given in Lifecodes' formal report to the district attorney

	D2S44	D17S79		DXYS14		
Blood from watch	10.25	3.87	3.50	4.83	3.00	1.94
Deceased mother	10.25	3.87	3.50	4.83	3.00	1.94
Deceased daughter	ND	3.87	3.50	4.83	—	1.94

five bands in the watch lane and only three bands in the mother's (Fig. 1). In his testimony, Michael Baird, Lifecodes' director of paternity and forensics, agreed that the watch lane showed two additional non-matching bands, but he asserted that these bands could be discounted as being contaminants "of a non-human origin that we have not been able to identify".

In my opinion, it is impossible to know whether the bands are non-human without demonstrating that they are absent when the experiment is repeated with an uncontaminated probe. How then did Lifecodes reach its judgment? Baird stated that, in his experience, DXYS14 should exhibit a pattern of fragments whose intensities decrease in proportion to their length: the extra bands could be ignored because their intensities were "not in the proportion I would expect to see".

In fact, the published scientific literature shows that DXYS14 actually yields patterns (ranging from one to more than six polymorphic fragments) whose intensities obey no ironclad rule. Ironically, the prosecution itself had put into evidence the very article which proved the point: the original paper<sup>2</sup> defining DXYS14, by Howard Cooke of the Medical Research Council in Edinburgh, containing photographs of *Pst*1 Southern blots showing lanes in which hybridization intensity and fragment length are clearly uncorrelated (Figs 6c and 7a in ref. 2).

David Page of the Whitehead Institute later introduced *Pst*1 blots from his laboratory showing arbitrary patterns of intensities at DXYS14. Finally, Cooke himself,

(For reasons I do not know, the forensic report listed only two of these bands.)

(2) Although D17S79 is expected to yield at most two bands, the dead daughter's lane exhibited four bands in the appropriate size range (not shown) in the only hybridization with the probe completed before the forensic report was issued. The report listed only two of these four bands — the two in the same position as in the mother and the watch.

The tendency to use lane-to-lane comparison to distinguish between bands and artefacts is perfectly natural; such comparison can be quite helpful in certain experiments. However, in my opinion, it is inappropriate in DNA fingerprinting analysis of unknown samples — as one runs the risk of discounting precisely those differences that would exonerate an innocent defendant. Forensic laboratories should be required to use objective criteria for identifying the bands in each lane, and to use experiments to rule out proposed artefacts.

When a result is reported to have an error rate of 1 in 100,000,000, it seems essential that the underlying data are not

left as a matter of subjective opinion.

**D2S44 and D17S79: declaring a match.** To obtain objective measurements of a band's position, Lifecodes uses a computer-digitizing apparatus<sup>3,5</sup>. The approach is reported to be highly accurate: when identical fragments are electrophoresed in different lanes, the difference between their positions is reported to show a standard deviation (s.d.) equal to 0.6 per cent of molecular weight<sup>3,5</sup>. Based on these experiments, Lifecodes defined a formal matching rule: two fragments are said to match when their positions differ by less than 3 s.d.s. The matching rule was explicitly stated in a recent population study<sup>5</sup> ("two DNA fragments were considered to be of different size if their values differed by more than 3 s.d.s") and in the formal forensic reports<sup>1</sup> in the Castro case ("fragments with measurements that are within [3 s.d.s.] of each other . . . are considered indistinguishable and their average size reported".)

Because the fragments at D2S44 and D17S79 did not appear to match perfectly, the defence examined Lifecodes' computer measurements. In fact, the bands fell outside the declared matching rule; as shown in Table 2, the bands at D2S44 differ by 3.06 s.d.s and the lower bands at D17S79 differ by 3.66 s.d.s. Under the objective matching rule, the bands were non-matches.

Why then was a match declared? Lifecodes stated that it did not actually use the objective threshold of 3 s.d.s. for declaring a forensic match: its decisions were based on subjective visual comparison. Agreeing that the explicit statements in the forensic report implied that the objective criterion had been used, Baird allowed that the statement "may not be the best explanation" of the company's actual procedures. As far as I can see, there is also no mention of the use of visual matching in the company's scientific papers or forensic reports. Clearly, there has been a significant misunderstanding about the matching rule which Lifecodes has been using.

In my opinion, visual matching is inappropriate in DNA fingerprinting, inasmuch as (1) many alleles have very similar sizes; (2) the accuracy of the measurement process is reported to be known; and (3) without an objective definition of a match, there is no meaningful

TABLE 2 Measured fragment sizes (in base pairs) at three RFLPs in *New York v. Castro*, as shown in Lifecodes' records from its computer-digitizing apparatus produced in response to subpoena

	DS244	D17S79		DXYS14		
Blood from watch	10,350	3,877	3,541	4,858	2,995	1,957
Deceased mother	10,162	3,869	3,464	4,855	2,999	1,946
Difference (per cent of average size)	1.83	0.21	2.20	0.06	0.13	0.56
Difference in number of s.d.s	3.06*	0.34	3.66*	0.10	0.22	0.94

According to Lifecodes' published papers<sup>3,5</sup> and its formal reports to district attorney<sup>1</sup>, the difference between two identical bands in different lanes shows an s.d. equal to 0.6 per cent of fragment size. Asterisks, bands differing by > 3 s.d.s.

way to determine the probability that a declared match might have arisen by chance (see below).

**DYZ1: Use of controls.** The DYZ1 locus provides a convenient method of identifying the sex of a sample: the locus is repeated about 2,000 times on the distal long arm of the Y chromosome, giving rise to an intense 3.7-kilobase (kb) *Pst*I band in males and no such band in females. Based on a hybridization with a probe for DYZ1, the blood on the watch was said to have come from a female.

Indeed, the mother, daughter and watch DNAs showed no male-specific band. But neither did the lane marked control. Who was the control?

(1) Initially, Baird testified that the control DNA came from the female-derived HeLa cell line.

(2) Two weeks later, however, the technician who actually performed the experiments testified that the control DNA came not from HeLa cells but rather, he recalled, from a male scientist.

(3) Baird then explained the absence of a positive signal in the now-male control lane by telling the court that the male scientist has a "short" Y chromosome which "does not react with this repeat sequence", a condition which is "fairly rare, but it does happen". In conventional genetic terminology, the individual was a genetic mutant deleted for the region.

(4) I then testified that the population frequency of such complete deletions is about 1 in  $10^3$ – $10^4$ , with most causing phenotypic abnormalities usually including complete sterility (D. Page, H. Cooke and K. Smith, personal communication). A normal male with such a deletion would be so rare as to be publishable.

(5) Baird then reported that the control DNA came not from the male scientist, but from a female technician. Although no precise record had been kept of which DNA preparation had been used, he said that he had managed to identify the source of the control lane by studying its RFLP pattern — an unforeseen use of DNA fingerprinting.

The confusion had probably resulted from faulty recollections (by Baird and the technician) and faulty inferences (about the male scientist), but it underscored the need for meticulous record-keeping in DNA forensics, which may not originally have been as clear.

Leaving aside the identity of the control DNA, there is a more important question: should a sex test be considered reliable without seeing a positive control on the autoradiogram to prove that the experiment had worked correctly? Baird testified that such a result could be considered "reliable". I would vigorously disagree.

**D2S44: Analysis of degraded DNA.** Based on seeing only a single band in the watch DNA, the sample was reported to be homozygous for a 10.25-kilobase (kb)



FIG. 1 Hybridization of probe 29C1 for the locus DXYS14 to *Pst*I-digested DNA from deceased mother (M), blood speck from defendant's watch (W) and deceased daughter (D), performed by Lifecodes in *New York v. Castro*. Although the autoradiogram shows three bands in lane M, five bands in lane W and one band in lane D, Lifecodes recorded three bands in identical positions in all three lanes. The autoradiogram is an overnight exposure; no further exposures or hybridizations involving this probe were performed.

band at the D2S44 locus. Although the conclusion may seem reasonable at face value, there is a serious problem: the small quantity of DNA on the watch was clearly degraded, and nearly 90 per cent of alleles in the Hispanic population lie above 10.25 kb.

How can one be sure that the sample was not a heterozygote, with a higher band undetected due to degradation? Estimating the extent of degradation by eye, Baird stated that the photographs of the ethidium bromide stain of the gel "gives you some indication that there is enough material present to be able to get a signal" in the 12–15-kb range", but that "I [d] hate to bet the ranch" on it.

Wouldn't a probe detecting a non-polymorphic single-copy band at about 15 kb have provided a definitive positive

control? He replied as follows:

"If you are making a decision based only on a single locus, whether or not a pattern you saw was homozygous or whether or not you are missing a band, you'd have to have some way to absolutely be sure that you were seeing everything that was there and the control that you mentioned would be very helpful to do that.

Now, in addition to the D2S44 locus, we also looked at two additional loci, D17S79 and DXYS14. By looking at the combination of loci, and seeing whether or not there was a pattern that matched or not, allowed us to determine or help[ed] us to determine that the pattern that we saw with the D2S44 locus is [a] homozygote."

Personally, I do not understand how the presence of matches at D17S79 and DXYS14 has any bearing on the determination of a match at D2S44: each test must be evaluated independently, especially as the individual probabilities of a match for each locus are multiplied together at the end (see below).

**Probe contamination.** To explain various artefacts, Lifecodes invoked four separate instances of probe contamination: human probes were said to be contaminated with bacterial sequences, and bacterial and plasmid probes were said to be contaminated with human sequences. Moreover, Baird testified that the company continued to use probes even after learning that they were contaminated, while apparently keeping no precise record of when such probes had been used.

Although the use of contaminated probes may be permissible in some types of experiment, it is, in my opinion, inappropriate in DNA forensics. Because the samples are of unknown origin and often contaminated, false matches can result but may be hard to recognize.

**Population genetics.** After declaring a forensic match, testing laboratories apply a three-step procedure to calculate the probability that the match might have arisen by chance in the population: (1) for each allele, one counts the frequency with which matching bands occur in a previously-drawn population sample (Lifecodes used a database<sup>7</sup> of US Hispanics in the Castro case); (2) for each locus, one then computes the probability of observing a matching genotype by applying the classical Hardy–Weinberg equations<sup>5,6</sup>, under the assumption that population is freely intermixing and thus contains no heterogeneous subgroups; and (3) for the complete RFLP pattern, one then multiplies the three single-locus probabilities, under the assumption that the genotypes at the loci are in linkage equilibrium (are uncorrelated). In fact, none of these steps stood up to scientific scrutiny in the Castro case.

**Probability of a match: inconsistent matching rules.** Whatever matching rule is used to declare a forensic match, it is axiomatic that the same matching rule must then be

used for counting the matches occurring in the population database. In fact, Lifecodes' calculations did not use the same matching rule.

■ To declare a forensic match with a given band, Lifecodes' published matching rule calls for examining a range of  $\pm 3$  s.d.s around the band. Obviously, the chance of a match arising at random is just the proportion of bands in this range.

■ To calculate the probability of a matching band occurring by chance, however, Lifecodes uses a completely different approach: the reported probability is essentially the frequency of bands occurring in a window of size  $\pm \frac{2}{3}$  s.d.s. More exactly, as described in an as-yet-unpublished paper<sup>8</sup>, it is a weighted-average of frequencies corresponding to such intervals, with the terms weighted according to a gaussian distribution centred on the estimated allele size.

Because the method involves averaging frequencies corresponding to intervals that are 4.5-fold smaller than those allowed for declaring a forensic match, the probability reported for each allele will typically be too small by a factor of about 4.5. Because each RFLP involves two alleles, the probability may thus be understated by a factor of about  $(4.5)^2$ , or about 20, for each locus. For a three-locus genotype, the error may thus be about 8,000-fold. (Of course, these calculations assume that one uses the 3 s.d. matching threshold. If the less stringent standard of visual matching is used, the discrepancy may be even greater.) Such a statistical procedure is like catching a match with a 10-foot-wide butterfly net, but then attempting to prove the difficulty of the feat by showing how hard it is to catch matches with a 6-inch-wide butterfly net.

How does Lifecodes justify its approach? Once a forensic match is declared between two bands, Lifecodes apparently considers the average fragment size to represent the allele present in the sample. It then estimates the population frequency of this allele, essentially by counting the bands within a range so narrow that it may not even include either of the two actual measurements. This approach substantially underestimates the true chance of a forensic match occurring at random, as it takes no account of the actual threshold used for declaring matches.

**Heterogeneity within the Hispanic population.** To justify applying the classical formulas of population genetics in the Castro case, the Hispanic population must be in Hardy-Weinberg equilibrium. In fact, Lifecodes' own data show that it is not. The classical test for Hardy-Weinberg equilibrium is based on Wahlund's principle<sup>7,9</sup> that the rate of homozygosity in a population containing distinct subgroups will be higher than would be expected under the assumption of random

mating. Applying this test to the Hispanic sample, one finds spectacular deviations from Hardy-Weinberg equilibrium: 17 per cent observed homozygotes at D2S44 and 13 per cent observed homozygotes at D17S79 compared with only 4 per cent expected at each locus, indicating, perhaps not surprisingly, the presence of genetically distinct subgroups within the Hispanic sample. (The expected 4 per cent frequency of homozygotes is based on the empirical probability of randomly drawing two alleles from the population sample that are either identical or so close together as to be scored as a single band; the minimum size difference needed to discriminate between one versus two bands in Lifecodes' experiments was stated explicitly in testimony and in a paper<sup>8</sup>.)

Once a population is known to be heterogeneous, one also cannot assume linkage equilibrium even for loci on different chromosomes: if an individual possesses an allele common among Puerto Ricans at one locus, it is more likely that he will do so at a second locus as well.

In fact, Lifecodes' population study<sup>2</sup> is scientifically valuable. From an evolutionary point of view, highly polymorphic loci contain many young alleles which may not be uniformly distributed within the Caucasian, Black or Hispanic populations. Studies of RFLP frequencies can reveal the detailed substructure of the human population, shedding light on migrations and trading patterns.

Such complexities, however, can undermine the use of simplified calculations in DNA forensics. Without the assumption of Hardy-Weinberg equilibrium and linkage equilibrium, there is no reliable way to convert allele frequencies into overall genotype frequencies: applying the classical equations can lead to spuriously low probabilities of a match. Possible solutions include empirical studies to identify ethnic subgroups that are in Hardy-Weinberg equilibrium and theoretical studies to derive appropriate correction factors for heterogeneity.

### The experts' statement

After stepping down from the witness stand in late April, I attended the next day a Cold Spring Harbor conference on genome mapping co-organized by Richard Roberts, who had been the prosecution's lead witness when the Castro case had begun in mid-February. As Roberts explained, his testimony had been intended simply to provide the court with a primer on DNA analysis. Concerned about the issues that had come to light, Roberts conceived a novel plan: a joint scientific meeting of the experts who had testified for either side to review the evidence.

At the meeting, held on 11 May, the experts agreed upon a consensus statement declaring that "the DNA data in this

case are not scientifically reliable enough to support the assertion that the samples . . . do or do not match. If these data were submitted to a peer-reviewed journal in support of a conclusion, they would not be accepted. Further experimentation would be required". In particular, the statement cited the inappropriateness of (1) discounting the extra bands at DXYS14; (2) declaring a match between bands at D2S44 and D17S79 whose measured positions differed by more than Lifecodes' announced threshold of 3 s.d.s; and (3) using a less-strict matching rule when declaring forensic matches than when searching the population database. (Baird was unable to attend the experts' meeting because of a prior engagement.)

At first, the prosecution indicated that it would withdraw the DNA evidence based on the advice of its own scientific experts. Eventually, however, the district attorney decided to press ahead.

The hearing then resumed, with former prosecution witnesses testifying now for the defence. The prosecution's efforts to mount a rebuttal case fizzled:

■ To rebut the non-matching bands at D2S44 and D17S79, the prosecution stated that the use of a new measurement system now reportedly showed that the bands actually lay within 3 s.d.s. The judge ruled this evidence inadmissible, however, calling a last-minute switch of methodology "highly unscientific".

■ To rebut the problem with degradation above 10 kb, Lifecodes probed the Southern blot with the human *Alu* repeat sequence and determined that it showed hybridization up to the 23-kb molecular mass marker. In my opinion, the experiment itself was meaningless (because the ability to detect a sequence repeated 300,000 times in the genome has no bearing on the ability to detect single-copy sequences), but it was unnecessary to explain this to the court. Defence attorney Peter Neufeld, by now a veteran reader of autoradiograms, noticed that someone had accidentally misread the size markers: the *Alu* hybridization actually extended only to the 9.8-kb marker.

■ To rebut the population-genetic issues, the prosecution made eleventh-hour phone calls to various scientists, but they refused to testify for the state.

In the end, the prosecution's rebuttal case consisted only of the contention that the DYZ1 sex-test hybridization had actually worked correctly — based on the fact that Lifecodes had located another experiment done on the same day which showed a male-specific band in a rape suspect. (This would be an adequate, if unorthodox, control if it could be proved that both hybridizations had been carried out in the same plastic bag.)

The hearing concluded on 26 May, about 15 weeks after it began. Lawyers for both sides are now preparing briefs. Jus-

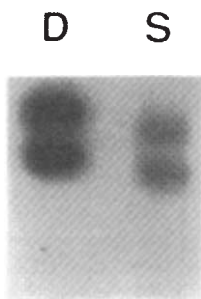


FIG. 2 Hybridization of probe pAC256 for the locus D17S79 to *Pst*I-digested DNA from defendant (D) and semen sample (S), performed by Lifecodes in *Georgia v. Caldwell*.

tice Gerald Sheindlin is expected to issue a decision in July. Sometime thereafter, the murder trial itself will commence — with or without the DNA evidence.

### Other illustrative cases

The general issues of interpretation are unique neither to the Castro case nor to Lifecodes, as the following examples show.

**Georgia v. Caldwell.** In a death-penalty case currently in progress in Atlanta, James Caldwell stands accused of raping and killing his daughter Sarah. According to the forensic report, Caldwell's blood matches semen samples from the crime with 67,500,000:1 odds against the match arising at random. Testifying at the Frye hearing, Lifecodes scientist Kevin McElfresh described the process of declaring a match as a "very simple straightforward operation", asserting that "there are no objective standards about making a visual match. Either it matches or it doesn't. It's like if you walk into the parking lot and see two blue Fords parked next to each other. That's the situation here."

In fact, the patterns clearly do not match by eye (Fig. 2). McElfresh agreed, but asserted that "There is, however, a consistent non-alignment of the bands throughout the test, telling us there's a match." In other words, McElfresh contended that the differences were due to one lane having run faster than the other, although Lifecodes presented no internal controls to support this explanation.

Although the hearing had been expected to last only a week, prosecutors asked for a month-long delay to prepare their rebuttal. Observers speculated that Lifecodes might use the time to test an internal control. If so, it would raise the question of whether Frye hearings are becoming a substitute for having generally accepted laboratory protocols in the first place.

**New York v. Neysmith.** In a Bronx case in 1987, Hamilton Neysmith was charged with rape based on the victim's identification. Asserting his innocence, Neysmith hired Lifecodes to compare his blood with semen samples from the assailant: the laboratory declared an exclusion. Protest-

ing that the defendant may not have sent his own blood, prosecutors obtained a court order to compel a second blood sample: Lifecodes reported in August 1988 that the two blood samples came from different people. Based on this evidence, Bronx assistant district attorney Karen Yaremko asked the court to revoke Neysmith's bail, planning to charge him with obstruction of justice. The judge declined to revoke bail and the defendant, rather than leaving town, maintained his innocence, and demanded a third blood sample. After Yaremko pressed Lifecodes, she said, the company determined that an error had indeed occurred. Having come close to losing his liberty over inaccurate DNA results, Neysmith was finally exonerated after blood and semen samples were sent to Cellmark Diagnostics (which confirmed the original exclusion. (Lifecodes declined last week to comment about the incident, which may have been nothing more than the sort of sample mix-up that can occur in any clinical laboratory. In view of the infallibility with which many jurors regard DNA fingerprinting, however, it may be that even stricter sample-handling procedures should be required.)

In *New York v. McNamara* in November 1988, another Bronx defendant sought to prove his innocence with DNA fingerprinting. Assistant district attorney Renee Myatt opposed the request, telling the court, "the office policy in dealing with a particular agency that does testing with respect to DNA [is] that their testing has been inaccurate, and therefore, unreliable." Notwithstanding this policy, the same District Attorney's office sought to introduce DNA evidence three months later in *New York v. Castro*.

**Texas v. Hicks.** In this rape-murder, the odds of the declared match occurring at random were reported to be 1 in 96,000,000. Apart from the issue of the matching rule used for searching the database, the population-genetic analysis took no account of the fact that the crime occurred in a small, inbred Texas town founded by a handful of families. The defendant was convicted and sentenced to death.

**The case of the abandoned baby.** When the president of an insurance agency in Ocean City, Maryland left her car to be towed for repairs to a garage in February 1988, the mechanic claimed that he had discovered a dead infant in the back seat. Although the woman insisted and a detailed medical examination confirmed that she had not been pregnant, Cellmark Diagnostics reported that its DNA analysis showed that she was the mother. (The evidence consisted of one hybridization with a mixture of four RFLP probes in which the woman shared four of eight bands with the child, as well as two hybridizations using probes detecting certain multi-locus

repeats.) Local papers reported the sensational news. No murder charges were filed, however, after the state medical examiner determined that the baby had been stillborn on about 4 February.

As it happens, the woman later gave birth to a full-term baby girl on October 24 — conceived on about January 29, according to sonograms carried out by her obstetrician. In view of the apparent contradiction, Cellmark last week invited a group of outside scientists to reanalyse remaining DNA samples from the baby.

### Setting standards

Readers should not conclude from this article that DNA fingerprinting is not a powerful tool for forensic identification or that current testing labs are not competent: as in the early stages of any new technology, some difficulties are to be expected. Rather, there is an urgent need for the scientific community to agree on clear guidelines for the procedures and standards needed to ensure reliable DNA fingerprinting. Legislators should also consider whether licensing and proficiency testing should be required in forensics. At present, forensic science is virtually unregulated — with the paradoxical result that clinical laboratories must meet higher standards to be allowed to diagnose strep throat than forensic labs must meet to put a defendant on death row.

An appropriate start would be a US National Academy of Sciences committee, charged with preparing a report on guidelines for DNA fingerprinting. There is ample precedent: when voice-print evidence began to be introduced in the 1970s, the academy convened such a group to examine the technology. An academy study on DNA fingerprinting had been planned for last year, but was postponed indefinitely when the National Institute of Justice would not finance it. As one justice official told me, the study was unwelcome: scientists had done their part by discovering DNA; it was not their job to tell forensic labs how to use it. □

*Eric S. Lander is in the Whitehead Institute for Biomedical Research, Nine Cambridge Center, Cambridge, Massachusetts 02142, USA.*

1. Lifecodes' formal reports to District Attorney (22 July 1987, 12 August 1988 and 6 February 1989).
2. Cooke, H. *Nature* **317**, 687–692 (1985).
3. Baird, M. et al. *Am. J. hum. Genet.* **39**, 489–501 (1986).
4. Baird, M. et al. *Adv. forens. Haemogenet.* **2**, 396–402 (1988).
5. Balasz, I., Baird, M., Clyne, M. & Meade, E. *Am. J. hum. Genet.* **44**, 182–190 (1989).
6. Cavalli-Sforza, L.L. & Bodmer, W.F. *The Genetics of Human Populations* (Sinauer, Sunderland, Mass., 1989).
7. Hartl, D. L. & Clark, A. G. *Principles of Population Genetics* (Sinauer, Sunderland, Massachusetts, 1989).
8. Morris, J.W., Sanda, A.I. & Glassberg, J.J. *forens. Sci.* (in the press).
9. Wahlund, S. *Hereditas* **11**, 65–106 (1928).

ACKNOWLEDGEMENTS. The Castro defence was a joint effort with my colleagues Howard Cooke, Lorraine Flaherty, Conrad Gilliam, Philip Green and David Page, together with attorneys Peter Neufeld and Barry Scheck. I also thank Richard Roberts for his efforts to help resolve this hearing and Carl Dobkin for participating in the experts' meeting. I am grateful to Susan Lee and Stephen Lincoln for technical assistance.